



INTER
FACES
CIENTÍFICAS

EDUCAÇÃO

ISSN IMPRESSO 2316-333X

E-ISSN 2316-3828

DOI-10.17564/2316-3828.2017v6n1p125-136

MINERAÇÃO DE DADOS EM AMBIENTES VIRTUAIS DE APRENDIZAGEM: APORTES PARA A PESQUISA EM EDUCAÇÃO A DISTÂNCIA

DATA MINING IN VIRTUAL LEARNING ENVIRONMENTS: CONTRIBUTIONS TO RESEARCH IN DISTANCE EDUCATION

MINERÍA DE DATOS EN ENTORNOS VIRTUALES DE APRENDIZAJE: APORTES PARA LA INVESTIGACIÓN EN EDUCACIÓN A DISTANCIA

Anabela Aparecida Silva Babosa¹

Rafael Nink de Carvalho³

Fábio Santos Andrade²

RESUMO

Na Educação a Distância existe um limbo no volume de registros sobre o processo de aprendizagem gerados pelo Ambiente Virtual de Aprendizagem. A mineração de dados serve para descobrir informações invisíveis por meio da análise de grandes quantidades de dados. Esse estudo investiga aspectos da mineração de dados como recurso tecnológico para as investigações no Ambiente Virtual de Aprendizagem dos cursos técnicos de Educação a Distância do Instituto Federal em Rondônia. A pesquisa apresenta os aportes e os procedimentos de pesquisa utilizando as técnicas de mineração de dados a partir da análise do banco

de dados e a ação reflexiva dos professores pesquisadores. Os resultados indicam que a mineração de dados permite acesso a informações sobre o processo educativo tais como aproveitamento, desempenho e predição do resultado final identificando relações entre dados que podem produzir novos conhecimentos sobre a realidade. Isso possibilita a reflexão sobre o ensinar e o aprender e por conseguinte a tomada de decisão com o uso destes dados para melhorar os resultados educacionais. A termo, sugere o aprofundamento sobre as contribuições deste processo para a pesquisa em educação.

PALAVRAS-CHAVE

Pesquisa. Educação a Distância. Mineração de Dados.

ABSTRACT

In Distance Education there is a limbo in the volume of records about the learning process generated by the Virtual Learning Environment. Data mining is for discovering invisible information by analyzing large amounts of data. This study investigates aspects of data mining as a technological resource for investigations in the Virtual Learning Environment of the Distance Education technical courses of the Federal Institute in Rondônia. The research presents the contributions and the research procedures using the techniques of data mining from the analysis of the database and the discussion with the teachers. The results indicate that data mining allows access to information about the educational process such as per-

formance, achievement and prediction of the final result, identifying relationships between data that can produce new knowledge about reality. This enables reflection on teaching and learning and therefore decision making using these data to improve educational outcomes. To conclude, it suggests the deepening on the contributions of this process for the research in education.

KEYWORDS

Research. Distance Education. Data Mining.

RESUMEN

En la educación a distancia existe un vacío en el volumen de registros sobre el proceso de aprendizaje generados por Entorno Virtual de Aprendizaje. La minería de datos se utiliza para descubrir información invisible a través del análisis de grandes cantidades de datos. Este estudio investiga los aspectos de la minería de datos como un recurso tecnológico para la investigación en el entorno de aprendizaje virtual de cursos técnicos de Educación a Distancia del Instituto Federal de Rondônia. La investigación presenta las contribuciones y los procedimientos de investigación utilizando técnicas de minería de datos a partir del análisis de bases de datos y la discusión con los maestros. Los resultados indican que la minería de datos permite el acceso a la información sobre el proceso educativo, tales como la

utilización, el rendimiento y la predicción del resultado final mediante la identificación de las relaciones entre los datos que se pueden producir nuevos conocimientos sobre la realidad. Esto permite la reflexión sobre la enseñanza y el aprendizaje y, por tanto, el proceso de toma con el uso de estos datos para mejorar los resultados educativos decisión. El término sugiere la profundización de las contribuciones de este proceso a la investigación en educación.

PALABRAS CLAVE

Investigación. Educación a distancia. Minería de datos.

1 INTRODUÇÃO

A cibercultura provoca mutações na educação. Na Educação a Distância (EAD) essa realidade tecnocêntrica se amplifica e se transforma com o uso de Ambientes Virtuais de Aprendizagem (AVA) que registra todos os dados, imortaliza as ações de ensino e aprendizagem, de modo que os fatos do ato educativo não se esvaem com o término do processo. Esses dados ficam armazenados no AVA e parece adequado indagar como investigar estes dados? Como realizar a pesquisa em educação diante desta tessitura?

Este artigo dispõe sobre as questões que envolvem a mineração de dados como ferramenta de pesquisa em educação a partir do AVA dos cursos técnicos de Educação a Distância ofertados pelo Instituto Federal de Educação, Ciência e Tecnologia de Rondônia (IFRO).

Para isso apresenta as ainda recentes disposições conceituais e usuais da mineração como técnica de análise de grandes quantidades de dados e instrumento de pesquisa em educação, bem como descreve o processo de mineração e a lógica de funcionamento que possibilita identificar dados e informações que possam subsidiar a tomada de decisão, sua capacidade para produzir novos conhecimentos e gerar novas descobertas científicas a partir da ressignificação pelos sujeitos envolvidos no processo educacional numa atitude de pesquisa-ação.

2 PREMISSAS CONCEITUAIS SOBRE MINERAÇÃO DE DADOS

Em uma realidade tecnológica as formas de pensar e fazer educação também passam a ser mediadas pela tecnologia. Essa realidade tecnológica tem sido considerada como cibercultura, que “especifica o conjunto de técnicas (materiais e intelectuais), de práticas, de atitudes, de modos de pensamento e de valores que se desenvolvem juntamente com o cres-

cimento do ciberespaço” (LÉVY, 2007, p. 17). A cibercultura traz em sua essência a digitalização, a virtualização, as redes e o volume de dados informacionais.

Na EaD a cibercultura é uma manifestação orgânica. Com o uso das Tecnologias da Informação e Comunicação (TIC) a EaD amplia sua capacidade, especialmente com a inserção de AVA, que registram todo o processo educativo, gerando dados exponenciais e informações, *a priori*, intangíveis ao professor.

A dificuldade de analisar a grande e diversa quantidade de dados armazenados no AVA e suas implicações é uma demanda necessária na EaD. Conhecer o processo educativo e identificar elementos que podem contribuir na regulação da aprendizagem depende de pesquisa que transforme os dados estáticos em conhecimento tangível, vivo. Os dados podem ser conhecimento vivo e subsidiar a intervenção pedagógica ou se tornarem inertes e servirem ao arquivo morto da ignorância.

Para que seja conhecimento vivo a atuação do professor é indispensável e o próprio uso da computação para qualificar o virtual é necessário, “uma solução promissora para o desenvolvimento dessas ferramentas é o uso de técnicas denominadas de Mineração de Dados Educacionais ou EDM (Educational Data Mining)” (GOTTARDO, 2012, p. 24).

O termo mineração de dados tem como objetivo descobrir novas informações por meio da análise de grandes quantidades de dados (WITTEN, 2008). Minerar dados é uma prática recente inclusive no campo das TIC,

The need to understand large, complex, information-rich data sets is common to virtually all fields of business, Science, and engineering. In the business world, corporate and customer data are becoming recognized as a strategic asset. The ability to extract useful knowledge hidden in data and to act on that knowledge is becoming increasingly important in today's competitive world. The entire process of applying a computer-based methodology, including new techniques, for discovering knowledge from data is called data mining. (KANTARDIZCK, 2002, p. 2).

Então, a mineração refere-se ao processo de identificar relações entre dados que podem produzir novos conhecimentos e gerar novas descobertas científicas, e por consequência subsidiarem novas ações, transformando a realidade posta. Então, a mineração de dados versa sobre

knowledge discovery in databases (KDD) is the automatic extraction of implicit and interesting patterns from large data collections. KDD can be used not only to learn the model for the learning process (Hamalainen, Suhonen, Sutinen, & Toivonen, 2004) or student modeling (Tang & McCalla, 2002) but also to evaluate and to improve e-learning systems (Zai"ane & Luo, 2001) by discovering useful learning information from learning portfolios (Hwang, Chang, & Chen, 2004). Web-based learning environments are able to record most learning behaviors of the students, and are hence able to provide a huge amount of learning profile. Recently, there is a growing interest in the automatic analysis of learner interaction data with web-based learning environments. (ROMERO, 2007, p. 135).

Portanto, consiste em identificar os dados registrados virtualmente que possam gerar alguma informação, esta informação a partir de um processo de (re)significação tem o potencial de tornar-se conhecimento importante sobre determinado processo ou realidade, que por sua vez pode ser utilizado para tomar decisão sobre esta realidade (QUILICI-GONZALEZ; ZAMPIROLI, 2014, p. 19-20), de modo a transformá-la.

Ora, este processo de (re)significação é complexo. Exige certa inquietação do sujeito, bem como este sujeito precisa saber a diferença entre dado, informação e conhecimento,

O problema da diferença entre informação e conhecimento. São níveis de realidade completamente diferentes. A sabedoria é reflexiva, que o conhecimento é organizador e que a informação se apresenta sob forma de unidades a rigor designáveis sob forma de bits. O que é importante? Não é a informação, é a computação que trata e, eu diria mesmo, que extrai informações do universo. Eu estou de acordo com Von Foerster ao dizer que as informações não existem na natureza. Nós as extraímos da natureza; nós transformamos os elementos e acontecimentos em signos, nós arrancamos a informação do ruído a partir das redundâncias (MORIN, 2015, p. 110).

Extrair os dados da realidade, tratá-los de modo a se apresentarem como informação e (re)significá-los para que se tornem conhecimento parece inerente a ação do sujeito e à educação na cibercultura.

As abordagens de mineração de dados para fins educativos apresentam-se crescente e ganham relevância pela capacidade de identificar os percursos e percalços do processo de ensino. Sob esta compreensão a utilização da mineração de dados na educação pode ser vista como uma técnica de investigação. De modo que as técnicas de mineração são aplicadas para conhecer o processo e transformá-lo em razão da qualidade.

Data mining techniques can discover useful information that can be used in formative evaluation to assist educators establish a pedagogical basis for decisions when designing or modifying an environment or teaching approach. The application of data mining in educational systems is an iterative cycle of hypothesis formation, testing, and refinement. (ROMERO, 2007, p. 136-137).

Assim, a mineração de dados pode servir para identificar diferentes informações sobre todo o processo educativo. Todavia na cibercultura não basta inserir as tecnologias no processo educativo, sem que estas proporcionem possibilidades ao aluno e ao professor na construção do conhecimento, conforme aponta Lévy (2007, p. 172):

Não se trata aqui de utilizar as tecnologias a qualquer custo, mas sim de acompanhar consciente e deliberadamente uma mudança de civilização que questiona profundamente as formas institucionais, as mentalidades e a cultura dos sistemas educacionais tradicionais e, sobretudo, os papéis de professor e de aluno.

Nessa tessitura, é condição da práxis educativa conceber e utilizar as tecnologias para pesquisa, análise de seus resultados e transformação da situação. No mesmo sentido, para que o AVA atenda ao modelo pedagógico concebido pela instituição ou intencionado pelos professores é necessário que os professores conheçam o ambiente, suas funcionalidades e potencialidades, tanto para propor estruturas de construção, quanto arquiteturas pedagógicas que promovam a construção do conhecimento.

Para tanto o professor deve direcionar a pesquisa neste espaço virtual. É o direcionamento da pesquisa que fundamentará a reflexão sobre o ensino, e, por conseguinte sobre a sua prática, para Freire (1987, p. 29),

Não há ensino sem pesquisa e pesquisa sem ensino. Enquanto ensino continuo buscando, reprocurando. Ensino porque busco, porque indaguei, porque indago e me indago. Pesquiso para constatar, constatando, intervenho, intervindo, educo e me educo. Pesquiso para “conhecer e o que ainda não conheço e comunicar ou anunciar a novidade”.

Em muitos casos, na EaD o professor não conhece a realidade contida nos dados ou como usá-los, implicando em apropriar-se da atitude de pesquisa e das ferramentas de mineração para encontrar o conhecimento invisível. Por isso, para que a mineração de dados se preste como pesquisa em educação eficiente e eficaz é imprescindível que haja a participação do professor.

Evidente, a proposta de mineração de dados como instrumento de pesquisa contempla a tríade ação-reflexão-ação e se coaduna com a pesquisa-ação uma vez que esta tem como ponto de partida a articulação entre a produção de conhecimentos para a conscientização dos sujeitos e solução de problemas socialmente significativos, como afirma Thiollent (2000).

Ora, nesta perspectiva verificar como a mineração de dados pode funcionar como instrumento de pesquisa sobre os dados advindos do AVA/IFRO, como podem ser tratados, que associações podem ser feitas, e diante das conclusões parciais que ações podem ser assumidas pelos agentes educativos quanto ao desempenho dos alunos dos cursos técnicos EaD merece investigação.

A quantidade de dados armazenados no AVA e sua aplicabilidade não podem ser simplesmente ignoradas. Para buscar esse conhecimento implícito dos fenômenos virtuais deve estar sustentado por uma teoria filosófica – pedagógica que o mine para a reflexão-ação-reflexão, como Morin (2015, p. 22):

As teorias não são objetivas, são subjetivas-objetivas: tratam dados objetivos mas são construções, sistemas de ideias que se encontram aplicados ao mundo real para

lhe detectar as estruturas invisíveis, uma vez que a ciência se interessa, não pelos fenômenos, o que é trivial, mas pelo que está escondido por trás dos fenômenos.

Ipsi facto, a teoria e a pesquisa são condição *sine quo noun* para procurar o que está por trás dos fenômenos e lhe atribuir valor, significado de modo que este sirva para alterar sua prática.

Como indica Pedro Demo (2003, p. 63) “ver a realidade a olho nu é uma coisa, claramente. Outra coisa, é vê-la com o auxílio de ferramentas tecnológicas, pois podemos obter resultados muito surpreendentes, como a sua tessitura atômica, invisível a olho nu”. É justamente este o propósito da mineração de dados, possibilitar ao professor uma ferramenta tecnológica que o auxilia a ver o invisível da realidade.

Deste modo, e em especial na EaD, a prática pedagógica não é uma ação isolada e a reflexão-ação consiste justamente a partir de uma base teórica crítica sobre os dados armazenados no AVA utilizar as tecnologias (mineração) como ferramenta de pesquisa para o desenvolvimento de estratégias de intervenção pedagógica na realidade educacional.

3 APONTAMENTOS SOBRE O PROCESSO E O MÉTODO DA PESQUISA A PARTIR DA MINERAÇÃO DE DADOS

Este estudo compreende um recorte da pesquisa registrada sob o protocolo número CAEE: 56618716.0.0000.5653 e aprovada pelo parecer número 1.764.140, sendo conduzida de acordo com os padrões éticos exigidos e primando pelo sigilo dos dados que poderiam identificar quaisquer envolvidos.

Os dados, objeto de estudo referem-se à base de dados gerada a partir da versão 2.6 do *MOODLE®*. Foram utilizadas duas versões, sendo a versão inicial correspondente aos dados do período de 26/02/2013 (implantação do AVA) até a data 14/03/2015, essa versão apresenta a base de dados completa, compreen-

dendo 382 tabelas e nestas armazenados um total de 22.610.203 registros e uma segunda versão com um total de 24.945.960 registro armazenados no período de 26/02/2013 a 20/12/2016 referente aos cursos técnicos EaD do IFRO/Campus Porto Velho Zona Norte que atende 26 municípios do estado de Rondônia.

É importante destacar as etapas da mineração de dados do ponto de vista operacional, que de acordo com Castro e Ferrari (2016, p.5-6) consiste basicamente em 4: Base de dados, Preparação ou pré-processamento de dados Mineração de dados e Avaliação ou validação do conhecimento.

Estas etapas da mineração devem primar pela produção do conhecimento, aquele que será utilizado para a tomada de decisão, para isso depende de algumas tarefas,

As tarefas de mineração são realizadas em descrições e predições e para que isso ocorra são construídos e utilizados algoritmos, são estes mecanismos que promovem “análises descritivas permitem uma sumarização e compreensão dos objetos da base e seus atributos” (CASTRO; FERRARI, 2016, p. 8).

Então para realizar a pesquisa a partir da mineração foi utilizado o software PG Admin 3 que permite um acesso visual ao banco, lista de tabela e seus campos, reconhecendo o que fica registrado do ensino-aprendizagem no AVA.

A partir dos dados organizados (dados do aluno, vínculos das disciplinas e polos, conceitos das atividades desenvolvidas, registro de frequência, acesso ao ambiente, interação com o tutor, entre outros) o processo de mineração realizou-se, utilizando técnicas de mineração por meio do software *Orange Data Mining* em sua versão 3.4.1. Este é responsável pela manipulação dos dados bem com os tratamentos estatísticos, e também pelas análises descritivas e de predição.

Para efetuar a mineração de dados foram necessárias a organização e a preparação da base de dados que se dão pela etapa de processamento, somente a partir do pré-processamento é possível minerar os dados. A mineração foi realizada em três etapas: pré-processamento, análise descritiva e análise preditiva por classificação de dados.

4 A PESQUISA A PARTIR DA MINERAÇÃO DE DADOS: OCORRÊNCIAS, EVIDÊNCIAS E EMERGÊNCIAS DO AVA

Este recorte apresenta a primeira fase da pesquisa no banco de dados que consistiu na análise da base de dados preexistente e reconhecimento dos dados armazenados extraíndo as ocorrências, evidências e emergências do AVA de acordo com as três etapas supramencionadas.

O pré-processamento estudou os dados armazenados nas tabelas que são estruturados e conexos entre as tabelas, utilizando chaves primárias e secundárias, as tabelas contêm linhas e colunas, sendo usual a nomenclatura de objeto ou instância para as linhas e atributos para as colunas das tabelas. Os atributos podem ser categóricos ou numéricos (CASTRO; FERRARI, 2016, p. 30).

Os dados estudados classificados como categóricos para efeito da mineração foram especificados como binário quando se tratava de valores lógicos (sim ou não) e os dados nominais quando armazenava valores literais (exemplo: nome do aluno, nome da disciplina, entre outros). Os dados classificados como numéricos podiam ser discretos quando armazenava valores numéricos inteiros, contínuos quando armazenava valores decimais e por último quando armazenava percentuais.

Na etapa de pré-processamento seguiu-se os procedimentos descritos por Castro e Ferrari (2016).

A definição de problema consistia em identificar como a utilização mineração de dados do AVA pode contribuir para melhorar o desempenho acadêmico, na sequência na seleção de dados brutos para a utilização da mineração foram estabelecidas algumas hipóteses iniciais:

- I. Hipótese 01: A disciplina ofertada influencia no resultado, ou seja, as disciplinas têm resultados próximos ou difere em graus de dificuldade;
- II. Hipótese 02: O quantitativo de interações do aluno no AVA está relacionado com o desempenho acadêmico;

III. Hipótese 03: O fator localidade interfere no desempenho acadêmico;

IV. Hipótese 04: Há possibilidade de prever o resultado final durante o processo.

Posterior às hipóteses definiu-se o algoritmo de pré-processamento da base, devido à complexidade da base de dados e o volume de informações contidas optou-se por reduzir a base de dados a duas tabelas, sendo uma para armazenar informações da disciplina e outra para agregar informações dos alunos quanto a essas disciplinas. Esse procedimento é justificado pela necessidade de múltiplas consultas inter-relacionadas o que exigia um enorme poder computacional.

A limpeza de dados é mínima devido à organização interna exigida pelo MOODLE® e isto ocorreu diretamente na etapa de mineração. Foi necessário fazer a integração dos dados já que a análise iria partir das duas bases existentes. A integração ocorreu na etapa de redução mesclando as informações. Não houve a necessidade de transformação porque os dados já estavam em uma base estruturada bem como a discretização, pois com exceção das notas os demais valores não eram contínuos, as notas finais que seriam utilizadas para análise sofreram um arredondamento para valores inteiros de 0 a 100. Os processos de discretização e normalização, quando necessário, ocorrem diretamente na etapa de mineração.

Diversas técnicas estatísticas são utilizadas na mineração como a análise descritiva, a modelagem preditiva, segmentação, entre outras. Fizeram-se mãos da análise descritiva e a modelagem preditiva assim definida

A análise descritiva de dados (ADD) é utilizada para descrever, simplificar ou sumarizar as principais características de uma base de dados, formando o princípio de qualquer análise quantitativa de dados. A diferença entre a análise descritiva e a mineração propriamente dita é que a ADD visa descrever e encontrar o que há nos dados, ao passo que os algoritmos de mineração buscam conclusões que extrapolam os dados e permitem inferir algo a partir deles. (CASTRO; FERRARI, 2016, p. 60).

Para a análise descritiva os dados são organizados e calculados as medidas de posição, dispersão e separatrizes. Para a modelagem preditiva os dados são classificados utilizando algoritmos específicos.

Conforme explicitado aqui na etapa de classificação o pesquisador pré-processa os dados e separa uma amostra para que o software de mineração utilize como treinamento, criando classificadores e com base nesses classificadores, a partir de outra amostra distinta teste a validade destes, certifique qual o método de classificação e apresente o melhor resultado. A saída é chamada de matriz de confusão que explicita os acertos e erros por categoria.

Pela análise descritiva foi possível detectar disciplinas com menor, ou melhor, aproveitamento, a dispersão da nota em função do desempenho se há o polo que apresenta resultados muito divergentes dos demais, entre outras indagações que ainda estão em trabalho. A partir dos dados organizados na etapa de pré-processamento e por meio de filtros e módulos do *software Orange* foi possível a determinação das medidas estatísticas.

Para identificar possíveis disciplinas que apresentam baixo desempenho de forma global, fato que poderia nortear o planejamento pedagógico para oferta da disciplina, como a escolha por não ofertar disciplina considerada mais exigente de forma simultânea, traçou o gráfico *boxplot* que permite uma visão quanto as medidas de posição, dispersão, assimetria, amplitude entre outros fatores.

Os dados gerados pelo gráfico englobam o desempenho nas atividades de percurso por disciplina referente à primeira turma do curso de Finanças (2013.1) e indicam que na disciplina de contabilidade de custo o desempenho nas atividades de percurso teve em média 45,7 pontos com desvio padrão de 22 pontos enquanto gestão tributária teve desempenho médio de 76,2 pontos com desvio padrão de 23 pontos.

Logo, estes dados podem gerar aos professores diversas indagações, como, por que o desempenho é tão diferente? O que levou ao baixo desempenho em contabilidade de custo? Será que as ferramentas utilizadas prejudicaram o andamento da atividade? Será que

a abordagem foi distinta do que a atividade proposta? Essas questões podem subsidiar o repensar da *práxis*.

De forma semelhante pode se observar o desempenho final por disciplina referente à primeira turma do curso de Finanças (2013.1).

A análise do resultado final apresentou menor dispersão em relação às atividades de percurso, porém na visualização destaca-se a disciplina de introdução à informática com desvio padrão de 23,6 pontos e matemática financeira com 23,8 pontos, ou seja, o desvio padrão que representa como as notas dos alunos se comportam ao longo da escala de 0 a 100 pontos é significativo, quanto mais alto for o desvio padrão maior será disperso das notas.

Do ponto de vista pedagógico significa que há um grupo com baixo desempenho, supostamente com dificuldades e outro com bom desempenho. No gráfico acima em introdução a informática, 25% dos alunos tiveram desempenho inferior a 48 pontos, outros 25% com desempenho entre 48 e 67 pontos, outros 25% com desempenho entre 67 e 77 pontos e finalmente 25% com desempenho superior a 77 pontos, o que demonstra o quão disperso estão os desempenhos.

Ainda quanto ao desempenho, mas relacionando com as interações com o ambiente foi elaborado pelo *software Orange* – Base de Dados/AVA/IFRO 2017 um gráfico que apresenta a relação entre desempenho e interações no AVA por disciplina referente à primeira turma do curso de Finanças (2013.1).

Para a relação entre nota e interações no AVA se baseou apenas no quantitativo e não na qualidade do acesso, quanto tempo ficou conectado ou que o aluno acessou e sim meramente a quantidade de interações. Foi possível observar que não há uma relação de proporcionalidade em que mais acessos garantem melhor desempenho, mas foi possível observar a partir da construção de uma tabela de contingência entre as variáveis média final e interações AVA que coeficiente¹ $\phi = 0,62$ ($N=7765$ e $X^2=3029,18$), esse valor varia de 0 que corresponde a ausência de associação

entre as variáveis e o valor 1 que representa uma associação completa, ou seja, o valor está mais próximo de uma associação do que a ausência de associação entre as variáveis. Então existe certa associação a ser feita ao grupo de alunos com maior interação no AVA que possui relativamente maior nota.

Ainda quanto ao desempenho, mas agora relacionando o seu comportamento frente à localidade, aos polos de oferta, dados referente à primeira turma do curso de Finanças (2013/1) quanto aos resultados finais.

Para fins desta pesquisa e para contornar as limitações do AVA e de registro acadêmico foram considerados os possíveis resultados: Aprovado - se o aluno conseguiu atingir nota superior ou igual a 60% e presença superior ou igual a 75%; Reprovado - se atingir nota inferior a 60% e presença superior ou igual a 75%; Reprovado por Falta - se não atingir presença superior ou igual a 75% dos momentos presenciais com pelo menos uma frequência, ou Evadido - caso não tenha participado de nenhum momento presencial.

Os dados explicitam que há polos com 58% de aprovação frente a outros que conseguiram 92% de aprovação, o que levanta a hipótese de ser um fator a localidade em que o aluno estuda.

Na análise preditiva é feita a partir de uma amostra utilizada como treinamento prever o valor de uma determinada variável para o conjunto de dados distintos, por exemplo, utilizando os dados referentes ao aproveitamento nas atividades de percursos, a frequência nas aulas presenciais e a quantidade de acessos² ao AVA é possível prever o resultado final do aluno. Para esse procedimento, foram realizadas as seguintes etapas:

- I. Seleção da base de dados;
- II. Pré-processamento – limpar possíveis registros que não contenham a nota da atividade de percurso, ou registro de frequência ou o número de acesso;
- III. Seleção das bases de treinamento (Turma de Finanças 20131) e seleção da base de testes (Turma

1. O coeficiente de correlação ϕ (Phi) $\phi = \sqrt{\frac{x^2}{N}}$ é dado por x^2 , onde é referente ao Quiquadrado de Pearson da Associação e N é o número total de casos.

2. Considerou acessos ao AVA qualquer interação com o sistema, a exemplo, abrir a página do AVA, fazer login, acessar a disciplina, abrir o fórum, responder um questionário etc.

de Finanças 2014.1);

IV. Aplicação dos algoritmos de mineração: Árvore de classificação, regressão logística, k-Vizinho mais Próximos (do inglês, kNN), regras de indução (classificador CN2), *Random forest*, *Naive Bayes*, entre outros;

V. A predição, confrontando a base de treinamento e testes para verificar a acurácia, ou seja, o percentual de acertos.

A partir da sequência de testes anteriores verificou-se que o método de predição com melhor desempenho foi o de árvore de classificação. O algoritmo de árvore cria uma estrutura encadeada de classificações baseadas no histórico dos dados de treinamento. Com base em uma matriz de confusão foi identificado que o algoritmo de classificação teve uma acurácia de aproximadamente 93%, ou seja, antes de corrigir a avaliação presencial e das etapas de recuperação e exame final o algoritmo conseguiu acertar aproximadamente 93% dos resultados finais.

Portanto pela análise descritiva determinou-se que as disciplinas ofertadas influenciam no resultado, pois possuem desempenhos diversos quanto ao comportamento médio das notas bem como sua dispersão. Quanto ao quantitativo de interações do aluno com o AVA a percepção é que existe uma associação média com o desempenho.

O fator localidade interfere no desempenho acadêmico uma vez que se identificam resultados finais díspares entre os polos, alguns com alta aprovação e outros com baixa. Pela análise preditiva, apenas baseando-se no registro de frequências, notas nas atividades de percurso e quantitativo de interações foi possível prever com uma acurácia de 93% o resultado final.

As hipóteses iniciais levantadas e atestadas comportam um relatório parcial dos dados minerados utilizado como provocação para levantamento de outras hipóteses. O processo de mineração de dados é cíclico e será aplicado novamente para validar as hipóteses a serem elencadas durante a análise reflexiva do professor.

De modo inacabado e ainda provisório indica-se que a mineração pode identificar dados importantes para a ação-reflexão pedagógica e consequentemente

ser e configurar-se como uma perspectiva teórico-metodológica de pesquisa em educação na cibercultura.

5 CONSIDERAÇÕES FINAIS

A investigação possibilita estabelecer a mineração de dados funciona como ferramenta de pesquisa em ambientes virtuais de aprendizagem na educação a distância. É possível indicar que a mineração estabelece uma nova relação com o saber na cibercultura, saindo do caótico e desordenado registro de dados para a produção do conhecimento por meio da assunção de uma atitude de pesquisa.

As técnicas de mineração promovem a identificação de informações como: fatores que causam impactos no processo de aprendizagem, grau de aproveitamento, reprovação e evasão, motivos condicionantes do baixo desempenho dos alunos, a frequência e interação e sua relação com o rendimento, e ainda a capacidade de predição do resultado final.

Os resultados da pesquisa indicam que a mineração de dados permite acesso a informações sobre o processo educativo o que possibilita a reflexão sobre o ensinar e o aprender e, por conseguinte a produção de conhecimento, isto conduz a percepção crítica da prática educativa.

A mineração pode servir a pesquisa em educação na cibercultura, suas técnicas tem o condão de identificar dados, e por meio da ressignificação do professor transformá-los em conhecimento capaz de subsidiar a concepção de estratégias: intervenção pedagógica e regulação da aprendizagem.

A pesquisa em educação a partir da mineração de dados oferece informações preciosas a serem aproveitadas não apenas pelo professor, mas também pela instituição como elemento de subsídio para a identificação de itens importantes no planejamento, execução, monitoramento e avaliação da aula, da disciplina e do curso. O conhecimento obtido por meio da mineração pode ser considerado no planejamento e construção dos ambientes virtuais de aprendizagem, na

proposta metodológica do curso, na abordagem pedagógica das ferramentas de aprendizagem e tem o viés de contribuir para minimizar problemas de estrutura nos cursos técnicos EaD do IFRO, colaborando para a melhoria da educação em todo o estado de Rondônia.

REFERÊNCIAS

CASTRO, Leandro Nunes de; FERRARI, Daniel Gomes. **Introdução à mineração de dados:** conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.

DEMO, Pedro. **Conhecimento e aprendizagem na nova mídia.** Brasília: Plano, 2003.

FREIRE, P. **Pedagogia do oprimido.** Rio de Janeiro: Paz e Terra, 1987.

GOTTARDO, E.; KAESTNER, C; NORONHA, R.V. Aplicação de técnicas de mineração de dados para estimativa de desempenho acadêmico de estudantes em um AVA utilizando dados com classes desbalanceadas. Instituto Federal de Educação Ciência e Tecnologia do Rio Grande do Sul (IFRS).

Kantardzic M., R. P. Pasic, C. Templeman, R. Levine, “Data Mining Approach in a Selection of Laparoscopic Techniques”, **Proceedings of the 10th International Conference on Intelligent Systems, Arlington, VA, U.S.A.**, June 2001, pp. 1–4. Google Scholar

Kantardzic, M., Data **Mining: Concepts, Models, Methods, and Algorithms**, IEEE Press & John Wiley, November 2002.

LÉVY, P. **Cibercultura.** Trad. Carlos Irineu da Costa. São Paulo: 34, 2007.

MORIN, E. **A via para o futuro da humanidade.** Tradução Edgard de Assis Carvalho. Rio de Janeiro: Bertrand Brasil, 2015.

QUILICI-GONZALEZ, José Arthur; ZAMPIROLI, Francisco de Assis. **Sistemas inteligentes e mineração de dados.** Santo André: Triunfal, 2014.

ROMERO, C.; VENTURA, S. Educational data mining: A survey from 1995 to 2005. **Expert Systems with Applications**, v.33, n.1, p.135-146, jul. 2007. Department of Computer Sciences, University of Cordoba, Cordoba, Spain, Elsevier Ltd. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.4774&rep=rep1&type=pdf>>. Acesso em: 10 dez. 2016.

THIOLLENT, M. **Metodologia da pesquisa-ação.** São Paulo: Cortez, 2000.

WITTEN, I.H., Frank, E. **Data mining:** Practical Machine Learning Tools and Techniques with Java Implementations. San Francisco-CA: Morgan Kaufmann, 2008.

Recebido em: 6 de Junho de 2017
Avaliado em: 7 de Junho de 2017
Aceito em: 3 de Agosto de 2017

1. Professora do Instituto Federal de Educação, Ciência e Tecnologia de Rondônia - IFRO. Graduada em Pedagogia. Especialista em Metodologia do Ensino Superior e Mestranda em Educação Escolar pela Universidade Federal de Rondônia- UNIR. E-mail: annabellabarbosa@gmail.com

2. Doutor em Educação pela Universidade Federal de Mato Grosso UFMT. Mestre em Ciências Sociais (área de Sociologia) pela Pontifícia Universidade Católica de São Paulo PUC. Educador Social. Professor da Universidade Federal de Rondônia - UNIR. E-mail: fasaan@hotmail.com

3. Mestre em Matemática pela Universidade Federal de Rondônia - UNIR. Professor do Instituto Federal de Educação, Ciência e Tecnologia de Rondônia - IFRO.. E-mail: rafaelnink@gmail.com